

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>A61K 38/51, C12P 21/06, C12N 9/12, 1/20, 15/00, C07H 21/04</b>		A1	(11) International Publication Number: <b>WO 97/37683</b> (43) International Publication Date: <b>16 October 1997 (16.10.97)</b>
(21) International Application Number: <b>PCT/US97/05968</b> (22) International Filing Date: <b>10 April 1997 (10.04.97)</b>		(81) Designated States: CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Priority Data: 60/015,241 10 April 1996 (10.04.96) 08/831,590 9 April 1997 (09.04.97)		US US	Published <i>With international search report.</i>
(71) Applicant: CYTEL CORPORATION [US/US]; 3525 John Hopkins Court, San Diego, CA 92121 (US).			
(72) Inventor: KETCHAM, Catherine, M.; 326 Melrose Avenue, Encinitas, CA 92024 (US).			
(74) Agent: TOWNSEND AND TOWNSEND AND CREW L.L.P.; 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).			

(54) Title: NUCLEIC ACIDS ENCODING GDP-FUCOSE PYROPHOSPHORYLASE

(57) Abstract

DNA sequences encoding GDP-fucose pyrophosphorylase (GTP + fucose-1-P → GDP-fucose + PP<sub>i</sub>), are provided. The enzymes can be used in the synthesis of desired carbohydrate structures.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MK	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KR	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

5

## NUCLEIC ACIDS ENCODING GDP-FUCOSE PYROPHOSPHORYLASE

### CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from U.S. provisional application serial  
10 No. 60/015,241, incorporated by reference. In addition, this application is related to  
copending application numbers 08/628,543 for "Improved Enzymatic Synthesis of  
Oligosaccharides" and 08/628,545 for "Improved Enzymatic Synthesis of Glycosidic  
Linkages", both filed on April 10, 1996, which are incorporated by reference.

15

### FIELD OF THE INVENTION

The present invention relates to isolated GDP-fucose pyrophosphorylase  
enzymes and nucleic acids encoding them. The enzymes are particularly useful in the  
synthesis of oligosaccharides.

20

### BACKGROUND OF THE INVENTION

Increased understanding of the role of carbohydrates as recognition  
elements on the surface of cells has led to increased interest in the production of  
carbohydrate molecules of defined structure. For instance, compounds comprising the  
sialyl Lewis ligands, sialyl Lewis<sup>X</sup> and sialyl Lewis<sup>A</sup> are present in leukocyte and non-  
leukocyte cell lines that bind to receptors such as the ELAM-1 and GMP 140 receptors.  
25 Polley *et al.*, *Proc. Natl. Acad. Sci., USA*, 88:6224 (1991) and Phillips *et al.*, *Science*,  
250:1130 (1990), *see, also*, U.S.S.N. 08/063,181.

Because of interest in making desired carbohydrate structures,  
30 glycosyltransferases and their role in enzyme-catalyzed synthesis of carbohydrates are  
presently being extensively studied. These enzymes exhibit high specificity and are  
useful in forming carbohydrate structures of defined sequence. Consequently,  
glycosyltransferases are increasingly employed as enzymatic catalysts in the synthesis of

a number of carbohydrates used for therapeutic and other purposes (Ito *et al.*, *Pure Appl. Chem.*, 65:753 (1993); U.S. Patents 5,352,670, and 5,374,541).

Synthesis of desired carbohydrate compounds has been achieved on preparative scales using enzymatic cycles using glycosyltransferases such as  $\beta$ 1,4 galactosyltransferase and  $\alpha$ 2,3 sialyltransferase (See, e.g., U.S. Patent No. 5,374,541; 5 WO 9425615; and Ichikawa, *et al.*, *J. Am. Chem. Soc.*, 114:9283-9298 (1992)).

Although fucosyltransferases have been cloned and expressed, enzymes for the production of guanosine 5-diphospho- $\beta$  L-fucose (GDP-fucose), the donor substrate for the fucosyltransferases are not readily available. Use of fucosyltransferase cycles 10 would be greatly facilitated if GDP-fucose can be readily regenerated enzymatically. The present invention fulfills these and other needs.

#### SUMMARY OF THE INVENTION

The present invention provides compositions comprising isolated 15 GDP-fucose pyrophosphorylase (GDPFPP) enzymes as well as nucleic acids that encode them.

The proteins of the invention can be derived from a variety of sources and specifically bind to antibodies generated against a protein having a sequence as shown in SEQ ID NO:2. A particular example of the proteins of the invention is provided: a 20 protein that has GDP-fucose pyrophosphorylase activity as defined herein, a molecular weight of approximately 66 kD, and a sequence as shown in SEQ. ID. No. 2.

The nucleic acids of the invention can be derived from a variety of sources and typically hybridize under stringent conditions to a nucleic acid with a sequence as 25 shown in SEQ ID NO:1. The nucleic acids may be incorporated into appropriate recombinant DNA construct comprising a promoter operably linked to the nucleic acid. The promoter can be selected to direct expression of in any desired cell, e.g., mammalian cells, insect cells, fungal cells, and the like.

The invention further comprises compositions and kits that contain (alone, or in combination): oligonucleotides that specifically hybridize to SEQ ID NO:1 under 30 stringent conditions, antibodies or antisera that specifically bind to SEQ ID NO:2, and a protein that specifically cross-reacts with SEQ ID NO:2 and has GDPFPP activity.

**Definitions**

As used herein, "GDP-fucose pyrophosphorylase" is an isolated polypeptide or protein preparation capable of catalyzing the following reaction: GTP + fucose I-P  $\rightarrow$  GDP-fucose + PP<sub>j</sub>. The term further refers to, and explicitly encompasses, proteins that specifically cross-react with specific antisera to SEQ ID NO:2.

A "conservative substitution", when describing a protein refers to a change in the amino acid composition of the protein that does not substantially alter the protein's activity. Thus, "conservatively modified variations" of a particular amino acid sequence refers to amino acid substitutions of those amino acids that are not critical for protein activity or substitution of amino acids with other amino acids having similar properties (e.g., acidic, basic, positively or negatively charged, polar or non-polar, etc.) such that the substitutions of even critical amino acids do not substantially alter activity. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

See also, Creighton (1984) Proteins W.H. Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations".

The term "isolated" is meant to refer to material which is substantially or essentially free from components which normally accompany the enzyme as found in its native state. Thus, the enzymes of the invention do not include materials normally associated with their *in situ* environment. Typically, isolated proteins of the invention are at least about 80% pure, usually at least about 90%, and preferably at least about 95% as measured by band intensity on a silver stained gel. The polypeptides may be purified from transgenic cells or cells which they are naturally expressed.

An example of the GDP-fucose pyrophosphorylase enzymes of the invention is a protein that has a molecular weight of about 66 kD and GDPFPP activity. GDPFPP enzymatic activity can be assayed in a number of ways. Three preferred assays are presented in detail below. An isolated, purified GDPFPP of the invention will 5 generally catalyze the production of GDP-fucose from GTP and fucose 1-phosphate under conditions described in assay # 2.

Two polynucleotides or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues in the two sequences is the same when aligned for maximum correspondence. Optimal alignment of sequences for comparison 10 may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics 15 Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection.

The term "substantial identity" means that a polypeptide comprises a sequence that has at least 80% sequence identity, preferably 90%, more preferably 95% or more, compared to a reference sequence over a comparison window of about 20 20 residues to about 600 residues-- typically about 50 to about 500 residues usually about 250 to 300 residues. The values of percent identity are determined using the programs above.

Another indication that polypeptide sequences are substantially identical is if one protein is immunologically reactive with antibodies raised against the other 25 protein. Thus, the GDPFPP enzymes of the invention include polypeptides immunologically reactive with antibodies raised against a protein having a sequence as shown in SEQ. ID. No. 2. The phrase "specifically binds to an antibody" or "specifically immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of 30 a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its

specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *ANTIBODIES, A LABORATORY MANUAL*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Stringent conditions

10 are sequence-dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5° to about 20° C, usually about 10° C to about 15° C, lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe.

15 Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60° C. For instance in a standard Southern hybridization procedure, stringent conditions will include an initial wash in 6X SSC at 42° C followed by one or more additional washes in 0.2X SSC at a temperature of at least about 55° C, typically about 60° C and often about 65° C.

20 Nucleotide sequences are also substantially identical for purposes of this application when the polypeptides which they encode are substantially identical. Thus, where one nucleic acid sequence encodes essentially the same polypeptide as a second nucleic acid sequence, the two nucleic acid sequences are substantially identical, even if they would not hybridize under stringent conditions due to silent substitutions permitted 25 by the genetic code (see, Darnell *et al.* (1990) *Molecular Cell Biology*, Second Edition *Scientific American Books* W.H. Freeman and Company New York for an explanation of codon degeneracy and the genetic code).

30 Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualization upon staining. For certain purposes high resolution will be needed and HPLC or a similar means for purification utilized.

The term "residue" refers to an amino acid (D or L) or amino acid mimetic incorporated in a oligopeptide by an amide bond or amide bond mimetic. An

amide bond mimetic of the invention includes peptide backbone modifications well known to those skilled in the art.

#### DETAILED DESCRIPTION OF THE INVENTION

5 The present invention provides isolated GDP-fucose pyrophosphorylase (GDPFPP) and nucleic acids encoding the enzyme. The nucleic acids are used to recombinantly express the enzyme, which can be used in a variety of applications. A particularly useful application is for regeneration of GDP-fucose in fucosyltransferase cycles in the synthesis of carbohydrates.

10 Little is known about the enzymes involved in the metabolism of GDP-fucose. In most organisms, GDP-fucose is formed from GDP-mannose by the sequential action of three distinct enzymatic activities: GDP-D-mannose 4,6-dehydratase, and a single protein that accomplishes epimerization and reduction, GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase (Ginsburg, *et al.*, *J. Biol. Chem.*, 236, 2389-2393 (1961) and Chang, *et al.*, *J. Biol. Chem.*, 263, 1693-1697 (1988)). A minor or "scavenge" pathway also exists, in which free fucose is phosphorylated by fucokinase to form fucose 1-phosphate, which, along with guanosine 5'-triphosphate (GTP), is used by GDP-fucose pyrophosphorylase to form GDP-fucose (Ginsburg, *et al.*, *J. Biol. Chem.*, 236, 2389-2393 (1961) and Reitman, *J. Biol. Chem.*, 255, 9900-9906 (1980)).

15 The enzymes of the invention are particularly useful in methods for the formation of a glycosidic linkage which takes place in a reaction medium comprising at least one glycosyl transferase (e.g., fucosyltransferase), a donor substrate (e.g., GDP-fucose) which is usually formed during the reaction by the action of the GDPFPP, an acceptor sugar and a divalent metal cation. The methods rely on the use of a glycosyl transferase to catalyze the addition of a saccharide to a substrate saccharide. The addition takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. Biomolecules as defined here include but are not limited to biologically significant molecules such as proteins (e.g., glycoproteins), and lipids (e.g., glycolipids, phospholipids, sphingolipids and gangliosides). In these methods, the divalent metal ion concentration is typically supplemented during the formation of the glycosidic linkage to replenish the concentration of the metal cation in the reaction medium between about 2 mM and about 75 mM.

The following abbreviations are used herein:

Ara	= arabinosyl;
Fru	= fructosyl;
Fuc	= fucosyl;
5 Gal	= galactosyl;
GalNAc	= N-acetylgalacto;
Glc	= glucosyl;
GlcNAc	= N-acetylgluco;
Man	= mannosyl; and
10 NeuAc	= sialyl (N-acetylneuraminy1).

Oligosaccharides are considered to have a reducing end and a non-reducing end, whether or not the saccharide at the reducing end is in fact a reducing sugar. In accordance with accepted nomenclature, oligosaccharides are depicted herein with the non-reducing end on the left and the reducing end on the right.

All oligosaccharides described herein are described with the name or abbreviation for the non-reducing saccharide (e.g., Gal), followed by the configuration of the glycosidic bond ( $\alpha$  or  $\beta$ ), the ring bond, the ring position of the reducing saccharide involved in the bond, and then the name or abbreviation of the reducing saccharide (e.g., GlcNAc). The linkage between two sugars may be expressed, for example, as 2,3, 2 $\rightarrow$ 3, or (2,3). Each saccharide is a pyranose.

#### A. General Methods

Much of the nomenclature and general laboratory procedures required in this application can be found in Sambrook, *et al.*, *MOLECULAR CLONING A LABORATORY MANUAL* (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989. The manual is hereinafter referred to as "Sambrook, *et al.*"

The practice of this invention involves the construction of recombinant nucleic acids and the expression of genes in transfected cells. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and

5 Kimmel, GUIDE TO MOLECULAR CLONING TECHNIQUES, *Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA (Berger); and CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1994

5 Supplement) (Ausubel).

10 Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q $\beta$ -replicase amplification and other RNA polymerase mediated techniques are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987)

10 U.S. Patent No. 4,683,202; PCR PROTOCOLS A GUIDE TO METHODS AND APPLICATIONS (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3, 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86, 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem* 35, 1826;

15 Landegren *et al.*, (1988) *Science* 241, 1077-1080; Van Brunt (1990) *Biotechnology* 8, 291-294; Wu and Wallace, (1989) *Gene* 4, 560; and Barringer *et al.* (1990) *Gene* 89, 117. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039.

20 The culture of cells used in the present invention, including cell lines and cultured cells from tissue or blood samples is well known in the art. Freshney (*Culture of Animal Cells, a Manual of Basic Technique, third edition* Wiley-Liss, New York (1994)) and the references cited therein provides a general guide to the culture of cells.

#### B. Methods for isolating DNA encoding GDPFPP

25 The nucleic acid compositions of this invention, whether RNA, cDNA, genomic DNA, or a hybrid of the various combinations, may be isolated from natural sources or may be synthesized *in vitro*. The nucleic acids claimed may be present in transformed or transfected whole cells, in a transformed or transfected cell lysate, or in a partially purified or substantially pure form.

30 Techniques for manipulation of nucleic acids encoding the enzymes of the invention, such as subcloning nucleic acid sequences encoding polypeptides into expression vectors, labelling probes, DNA hybridization, and the like are described generally in Sambrook *et al.*

Recombinant DNA techniques can be used to produce GDPFPP. In general, the DNA encoding the enzymes are first cloned or isolated in a form suitable for ligation into an expression vector. After ligation, the vectors containing the DNA fragments or inserts are introduced into a suitable host cell for expression of the recombinant enzymes. The enzymes are then isolated from the host cells.

There are various methods of isolating the DNA sequences encoding GDPFPP. Typically, the DNA is isolated from a genomic or cDNA library using labelled oligonucleotide probes specific for sequences in the DNA. The DNA libraries can be prepared from any organism or cell type that produces the enzymes. For instance, both prokaryotic and eukaryotic organisms can be used. Typically the nucleic acids are prepared from mammalian cells or tissue. The particular species from which the nucleic acids are isolated is not critical so long as enzymes of sufficient activity are encoded. Suitable sources include human, rat, mouse, rabbit, pig, and other mammals. Other organisms that can be used include *Caenorhabditis elegans*, the slime mold *Dictyostelium discoideum*, plants such as *Zea mays*, *Arabidopsis thaliana*, *Vigna radiata* (mung bean), algae such as *Fucus gardnieri*, bacteria such as but not limited to: *Mycobacterium kansasii*, *Mycobacterium szulgai*, *Mycobacterium avium*, and *Mycobacterium tuberculosis*, *Neisseria gonorrhoeae*, *Neisseria Meningitidis*, and *Neisseria lactamica*, *Helicobacter pylori*, *Streptococcus pyogenes*, the fungi: *Candida albicans*, *Schizosaccharomyces pombe*, *Aspergillus niger*, *Aspergillus nidulans*, the fruit fly *Drosophila melanogaster*, amphibians such as *Xenopus laevis*, etc. The sequence of the probes used to isolate the nucleic acids is based on the nucleic acid sequences disclosed here. If cDNA libraries are used, mRNA from kidney or thyroid tissue is typically used to prepare the cDNA (However, GDPFPP mRNA is ubiquitously expressed in all tissues assayed by Northern blot using a human cDNA probe, and any tissue that express the mRNA may be used).

The polymerase chain reaction can also be used to prepare the DNA. Polymerase chain reaction technology (PCR) is used to amplify GDPFPP nucleic acid sequences directly from mRNA, from cDNA, and from genomic libraries or cDNA libraries.

Appropriate primers and probes for amplifying the GDPFPP DNA's are generated from analysis of the DNA sequences. In brief, oligonucleotide primers that are complementary to the DNA sequences located on both sides of the region to be

amplified are synthesized. The polymerase chain reaction is then carried out using the two primers. See PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990). Primers can be selected to amplify the entire gene or to amplify smaller segments, as desired. Preferred probes for this purpose are 5'-TCA-GAT-ATC-GGG-GCT-ATG-GCA-GCT-GCT-AG-3' and 5'-ATA-GAT-ATC-TCT-GGA-ATG-TTA-CTC-AAA-AAG-GCA-A-3', respectively.

Oligonucleotides for use as probes and primers are chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage, S.L. and Caruthers, M.H., (1981) *Tetrahedron Lett.*, 22(20):1859-1862 using an automated synthesizer, as described in Needham-VanDevanter, D.R., *et al.* 1984, *Nucleic Acids Res.*, 12:6159-6168. Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson, J.D. and Regnier, F.E. (1983), *J. Chrom.*, 255:137-149.

The sequence of the synthetic oligonucleotides can be verified using the chemical degradation method of Maxam, A.M. and Gilbert, 1980, in W., Grossman, L. and Moldave, D., eds. Academic Press, New York, *Methods in Enzymology*, 65:499-560.

Other methods known to those of skill in the art may also be used to isolate DNA encoding all or part of the GDPFPP enzyme. See Sambrook, *et al.*

#### C. Expression of GDPFPP

Once desired DNAs are isolated and cloned, one may express the desired polypeptides in a recombinantly engineered cell such as bacteria, fungus (including yeast), insect (especially employing baculoviral vectors), and mammalian cells. It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of the DNA encoding the enzymes. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of natural or synthetic nucleic acids encoding GDPFPP will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and

translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding the enzymes. To obtain high level expression of a cloned gene, it is desirable to construct expression plasmids which contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for 5 translational initiation, and a transcription/translation terminator.

### 1. Expression in Prokaryotes

Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway as described by Yanofsky, C., 1984, *J. Bacteriol.*, 158:1018-1024 and the leftward promoter of phage lambda ( $P_L$ ) as described by Herskowitz, I. and Hagen, D., 1980, *Ann. Rev. Genet.*, 10:399-445. The inclusion of selection markers in DNA vectors transformed in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol. See Sambrook *et al.* for details concerning selection markers for use in *E. coli*.

15 The vector is selected to allow introduction into the appropriate host cell. Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA.

20 Expression systems for expressing the enzymes are available using *E. coli*, *Bacillus* sp. (Palva, I. *et al.*, 1983, *Gene* 22:229-235; Mosbach, K. *et al.*, *Nature* 302:543-545 and *Salmonella*. *E. coli* systems are preferred.

25 The polypeptides produced by prokaryote cells may not necessarily fold properly. During purification from *E. coli*, the expressed polypeptides may first be denatured and then renatured. This can be accomplished by solubilizing the bacterially produced proteins in a chaotropic agent such as guanidine HCl and reducing all the cysteine residues with a reducing agent such as beta-mercaptoethanol. The polypeptides are then renatured, either by slow dialysis or by gel filtration. U.S. Patent No. 4,511,503.

30 Detection of the expressed enzyme is achieved by methods known in the art as radioimmunoassays, Western blotting techniques, immunoprecipitation, or activity assays. Purification from *E. coli* can be achieved following procedures described in U.S. Patent No. 4,511,503.

## 2. Expression in Eukaryotes

A variety of eukaryotic expression systems such fungal cells (particularly yeast), insect cell lines and mammalian cells, are known to those of skill in the art. As explained briefly below, the enzymes may also be expressed in these eukaryotic systems.

5

### a. Expression in Fungal Cells, Including Yeast

In addition to use in recombinant expression systems, the isolated GDPFPP DNA sequences can also be expressed in fungal cells.

Techniques for transforming fungi are well known in the literature, and have been described, for instance, by Beggs, Hinnen *et al.* (*Proc. Natl. Acad. Sci. USA* 10 75: 1929-1933 (1978)), Yelton *et al.* (*Proc. Natl. Acad. Sci. USA* 81: 1740-1747 (1984)), and Russell (*Nature* 301: 167-169 (1983)). Typically, filamentous fungi, for example, strains of the fungi *Aspergillus* (McKnight *et al.*, U.S. Patent No. 4,935,349) are used. Examples of useful promoters include those derived from *Aspergillus nidulans* glycolytic genes, such as the *ADH3* promoter (McKnight *et al.*, *EMBO J.* 4: 2093-2099 (1985)) and the *tpiA* promoter. An example of a suitable terminator is the *ADH3* terminator (McKnight *et al.*). The expression units utilizing such components are cloned into vectors that are capable of insertion into the chromosomal DNA of *Aspergillus*.

Synthesis of heterologous proteins in yeast is well known and described in the literature. *Methods in Yeast Genetics*, Sherman, F., *et al.*, Cold Spring Harbor 20 Laboratory, (1982) is a well recognized work describing the various methods available to produce the enzymes in yeast.

Examples of promoters for use in yeast include GAL1,10 (Johnson, M., and Davies, R.W., 1984, *Mol. and Cell. Biol.*, 4:1440-1448) ADH2 (Russell, D., *et al.* 25 1983, *J. Biol. Chem.*, 258:2674-2682), PH05 (EMBO J. 6:675-680, 1982), and MFα1 (Herskowitz, I. and Oshima, Y., 1982, in THE MOLECULAR BIOLOGY OF THE YEAST SACCHAROMYCES, (eds. Strathern, J.N. Jones, E.W., and Broach, J.R., Cold Spring Harbor Lab., Cold Spring Harbor, N.Y., pp. 181-209. A multicopy plasmid with a selective marker such as Leu-2, URA-3, Trp-1, and His-3 is also desirable.

A number of yeast expression plasmids like YEp6, YEp13, YEp4 can be 30 used as vectors. A gene of interest can be fused to any of the promoters in various yeast vectors. The above-mentioned plasmids have been fully described in the literature (Botstein, *et al.*, 1979, *Gene*, 8:17-24; Broach, *et al.*, 1979, *Gene*, 8:121-133).

Two procedures are used in transforming yeast cells. In one case, yeast cells are first converted into protoplasts using zymolyase, lyticase or glusulase, followed by addition of DNA and polyethylene glycol (PEG). The PEG-treated protoplasts are then regenerated in a 3% agar medium under selective conditions. Details of this procedure are given in the papers by J.D. Beggs, 1978, *Nature* (London), 275:104-109; and Hinnen, A., *et al.*, 1978, *Proc. Natl. Acad. Sci. USA*, 75:1929-1933. The second procedure does not involve removal of the cell wall. Instead the cells are treated with lithium chloride or acetate and PEG and put on selective plates (Ito, H., *et al.*, 1983, *J. Bact.*, 153:163-168).

The enzymes can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysates. The monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassays or other standard immunoassay techniques.

**b. Expression in Mammalian and Insect Cell Cultures**

Illustrative of cell cultures useful for the production of GDPFPP are cells of insect or mammalian origin. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. Illustrative examples of mammalian cell lines include VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, W138, BHK, Cos-7 or MDCK cell lines.

As indicated above, the vector, *e. g.*, a plasmid, which is used to transform the host cell, preferably contains DNA sequences to initiate transcription and sequences to control the translation of the gene sequence. These sequences are referred to as expression control sequences. When the host cell is of insect or mammalian origin illustrative expression control sequences are obtained from the SV-40 promoter (*Science*, 222:524-527, 1983), the CMV 1.E. Promoter (*Proc. Natl. Acad. Sci.* 81:659-663, 1984) or the metallothionein promoter (*Nature* 296:39-42, 1982). The cloning vector containing the expression control sequences is cleaved using restriction enzymes and adjusted in size as necessary or desirable and ligated with DNA coding for desired polypeptides by means well known in the art.

As with yeast, when higher animal host cells are employed, polyadenylation or transcription terminator sequences from known mammalian genes need to be incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate

splicing of the transcript may also be included. An example of a splicing sequence is the VPI intron from SV40 (Sprague, J. *et al.*, 1983, *J. Virol.* 45: 773-781).

Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors.

5 Saveria-Campo, M., 1985, "Bovine Papilloma virus DNA a Eukaryotic Cloning Vector" in DNA Cloning Vol. II a Practical Approach Ed. D.M. Glover, IRL Press, Arlington, Virginia pp. 213-238.

10 The host cells are competent or rendered competent for transformation by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate precipitation, fusion of the recipient cells with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextran, electroporation and micro-injection of the DNA directly into the cells.

15 The transformed cells are cultured by means well known in the art.

15 Biochemical Methods in Cell Culture and Virology, Kuchler, R.J., Dowden, Hutchinson and Ross, Inc., (1977). The expressed enzymes are isolated from cells grown as suspensions or as monolayers. The latter are recovered by well known mechanical, chemical or enzymatic means.

20 D. Preparation of Antibodies to GDPFPP

Methods of producing polyclonal and monoclonal antibodies are known to those of skill in the art. *See, e.g.*, Coligan (1991), CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY; and Harlow and Lane (1989), ANTIBODIES: A LABORATORY MANUAL, Cold Spring Harbor Press, NY; Stites *et al.* (eds.) BASIC AND CLINICAL IMMUNOLOGY (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding (1986), MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2d ed.) Academic Press, New York, NY; and Kohler and Milstein (1975), *Nature*, 256:495-497. Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors. *See, 25 Huse *et al.* (1989), Science, 246:1275-1281; and Ward *et al.* (1989) Nature, 303:544-546.* The antibody can also be selected from a phage display library screened against GDPFPP (*see, e.g.* Vaughan *et al.* (1996) *Nature Biotechnology*, 14: 309-314 and references therein). The antibodies or antisera may be used to specifically

characterize, detect or isolate proteins that crossreact with SEQ ID NO:2, by any of a variety of methods: Western blotting, immunoaffinity chromatography, ELISA, immunofluorescence microscopy, etc.

For example, in order to produce antisera for use in an immunoassay, the 5 polypeptide of SEQ ID NO: 1 or a fragment thereof is isolated as described herein. For example, recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is immunized with the protein of SEQ ID NO: 2 or a peptide thereof, using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein 10 and conjugated to a carrier protein can be used an immunogen. Polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of  $10^4$  or greater are selected and tested for their cross 15 reactivity against non-GDPFPP or even GDPFPP from other celltypes or species or a peptide fragment thereof, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a  $K_D$  of at least about 0.1 mM, more usually at least about 1  $\mu$ M, preferably at least about .1  $\mu$ M or better, and most preferably, .01  $\mu$ M or better.

Specific binding to an antibody under such conditions may require an 20 antibody that is selected for its specificity for a particular protein. For example, antisera raised to a GDPFPP having the amino acid sequence depicted in Seq. ID No. 2 can be selected to obtain antibodies specifically immunoreactive with GDPFPP and not with other proteins.

A variety of immunoassay formats may be used to select antibodies 25 specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

30 Immunoassays in the competitive binding format can be used for the crossreactivity determinations. For example, the protein of Seq. ID No. 2 can be immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete

with the binding of the antisera to the immobilized protein is compared to the protein of Seq. ID No. 2. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are then 5 removed from the pooled antisera by immunoabsorption with the above-listed proteins.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein to the immunogen protein (*i.e.*, the protein of Seq. ID No. 2). In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each 10 protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than 10 times the amount of the protein of Seq. ID No. 2 that is required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the protein of Seq. ID No. 2.

15 This invention also embraces kits for detecting the presence of @ in tissue or blood samples which comprise a container containing antibodies selectively immunoreactive to the protein and instructional material for performing the test. The kit may also contain other components such as @, controls, buffer solutions, and secondary 20 antibodies. Kits for detecting antibodies to @ comprise a container containing an @ instructional material and may comprise other materials such as secondary antibodies and labels as described herein.

#### E. Methods for isolating and purifying GDPFPP

25 The enzymes of the invention can be directly prepared from cells or tissue in which the enzymes are naturally expressed, or from genetically engineered cells in which they are expressed.

Any of the organisms suitable for isolation of the nucleic acids of the invention can be used as source material for the enzyme. In the case of mammals, the protocol for isolation of the enzymes is generally as follows:

- 30 1. Obtain GDPFPP-expressing tissue.
2. Homogenize (*e.g.*, by sonication, or in a dounce homogenizer, or in a blender) in about 1-10 mls of homogenizing solution (containing protease inhibitors, preferably at about 4° C) per gram tissue.

3. Remove insoluble and particulate matter, by filtration and/or centrifugation.

4. The polypeptides of this invention may be purified to substantial purity by standard techniques well known in the art, including: selective precipitation in a solution in which the GDPFPP is relatively insoluble (e.g., ammonium sulfate, PEG); immunoaffinity chromatography using antibodies to a known GDPFPP such as SEQ ID NO:2; ion exchange chromatography on a commercially available or known ion exchange matrix (for example, DEAE DE-52); hydrophobic chromatography on a commercially available or known matrix (for example, Phenyl Sepharose); gel filtration on a commercially available or known sieving matrix (for example, S-300); dye-ligand chromatography (for example, using Yellow Sepharose and Red Sepharose); chromatofocusing (for example, using PBE 94). See, for instance, R. Scopes, **PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE**, Springer-Verlag: New York (1982).

The presence, amount and relative activity of GDPFPP in a given sample, especially following a given purification step, is monitored by performing a GDPFPP assay and a protein assay of an aliquot (e.g., 1-1,000  $\mu$ l) from the sample, as described herein. It is also monitored by sodium dodecyl sulfate polyacrylamide chromatography (SDS-PAGE). The optimal combination and order of purification steps may vary from tissue to tissue and from tissue donor species to species. Using the following example as a basic framework for a purification and analytical scheme, a skilled routineer is capable of determining the better combination and order of purification steps for a given tissue and tissue donor species.

GDPFPP produced by recombinant DNA technology may also be isolated and purified by standard techniques well known to those of skill in the art. Recombinantly produced polypeptides can be directly expressed or expressed as a fusion protein. The protein is then purified by a combination of cell lysis (e. g., sonication) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme release the desired enzyme, which can be purified as described above.

## EXAMPLES

## Example 1

Purification of GDPFPP from Porcine Kidney

## 5 1. Purification of Enzyme.

All purification steps were carried out at 4° C. All concentration steps were carried out with an Amicon concentrator with a Diaflo membrane, 10,000 molecular weight (MW) cut-off. A typical isolation protocol is described in detail below.

10 **Homogenization and Centrifugation:** One fresh porcine kidney (130 g) was chopped and weighed. Buffer A (10 mM Tris-HCl, pH 7.8, 1 mM ethylenediamine-tetraacetic acid [EDTA], 1 mM  $\beta$ -mercaptoethanol [ $\beta$ ME], 50 mM sucrose), was added at 2.5 ml per gram tissue and the tissue was homogenized in a blender for four 1 min. bursts. The sample was centrifuged for 40 min at 12,000 x g in a Beckman J-21 centrifuge. The supernatant fraction was filtered through cheesecloth 15 and ultracentrifuged at 100,000 x g, for 45 min.

15 **Ammonium Sulfate Precipitation:** A 0-30% ammonium sulfate cut was performed on the above supernatant fraction and the sample centrifuged for 40 min at 12,000 xg in a Beckman J-21 centrifuge. A 30-60% ammonium sulfate cut was then performed, and after centrifugation as before, the pellet was collected. It was 20 resuspended in and dialyzed against buffer A.

**PEG Precipitation:** To each 100 ml of dialyzed, resuspended ammonium sulfate pellet, 25 g PEG 8000 were added. The sample was centrifuged at 12,000 x g as above, the supernatant fraction discarded, and the pellet resuspended in buffer A.

25 **Ion Exchange Chromatography on DEAE DE-52:** A 3 cm x 30 cm DE-52 column was equilibrated in buffer A. After loading the sample the column was washed with the same buffer. The enzyme was eluted with a 0-200 mM NaCl gradient in buffer A. Active fractions were pooled.

30 **Hydrophobic Chromatography on Phenyl Sepharose:** Phenyl Sepharose (about 100 ml) was equilibrated in buffer A containing 1M ammonium sulfate. Ammonium sulfate was also added to 1M to the enzyme sample. After loading the sample, the column was washed with equilibration buffer. The enzyme was eluted from the column with a 1.0-0 M ammonium sulfate gradient. Active fractions were pooled.

**Gel Filtration on S-300:** A 1.6 x 120 cm column of S-300 was equilibrated with buffer A. The enzyme sample was concentrated to 1 ml before loading. Elution was performed with buffer A. Active fractions were pooled.

**Dye-Ligand Chromatography on Yellow Sepharose and Red Sepharose:**

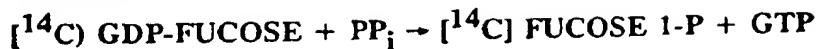
5 Sigma Reactive Yellow 86 Sepharose (approximately 20 ml) was equilibrated in buffer A. The pooled enzyme from the previous step was applied to the column. The column was washed with buffer A: runthrough and wash fractions, which contained the enzyme, were pooled and concentrated. (The bound contaminants were eluted from the resin with 2 M NaCl and discarded.) Sigma Reactive Red 120 Agarose (approximately 100 ml) 10 was also equilibrated with buffer A. The pooled enzyme that ran through the Yellow Sepharose was loaded and the column washed with buffer A. Some impurities were eluted with 5 mM ATP in buffer A. The enzyme was then eluted with 3 mM pyrophosphate in buffer A.

15 **Gel Filtration on Sephadryl S-300:** A 1.6 X 120 cm column of S-300 was equilibrated with buffer A. The enzyme sample was concentrated to 1 ml before loading. Elution was performed with buffer A.

20 **Chromatofocusing on PBE 94:** The PBE 94 column (approximately 20 ml) was equilibrated in 25 mM imidazole-HCl, pH 7.4, 50 mM sucrose, 1 mM EDTA, 1 mM  $\beta$ ME. The sample contain enzyme was concentrated to 1 ml and loaded. The column was washed with the equilibration buffer. The column was eluted with a self-forming pH gradient using Polybuffer 74-HCl, pH 4.2, diluted 1:8.

**2. Assays for GDPFPP Activity.**

25 Three assays for GDP-fucose pyrophosphorylase were used in these studies. Since the reaction is readily reversible, it is possible to assay for the synthesis or hydrolysis of GDP-fucose. Assay #1 measures the pyrophosphate (PP<sub>i</sub>)-dependent hydrolysis of radiolabeled GDP-fucose. This is the quicker and easier of the two assays and was used to monitor the purification. Assay #2 measures the synthesis of radiolabeled GDP-fucose from radiolabeled GTP and fucose 1-P. This assay is 30 time-consuming but is more specific, and was used to ensure that each completed preparation could indeed be used for the synthesis of GDP-fucose. Only assay #2 was used to screen the cells expressing the cloned enzyme.

**Assay #1.**

The following reagents are mixed together:

**Final concentration:**

5 100 mM Tris-HCl, pH 7.5

5 mM sodium pyrophosphate

4 mM MgCl<sub>2</sub>

5000 cpm [<sup>14</sup>C] GDP-fucose

diluted in 200 mM Tris-HCl,

10 pH 7.5

10  $\mu\text{l}$  enzyme (or enzyme diluted to 10  $\mu\text{l}$  in 200 mM Tris-HCl, pH 7.5), in a reaction volume of 50  $\mu\text{l}$ .

15 Incubate 5-10 min, 37° C.

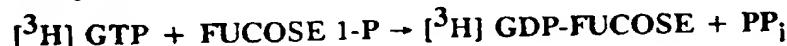
Add 500  $\mu\text{l}$  5% (w/v) trichloroacetic acid. Vortex.

Add 300  $\mu\text{l}$  Darco G-50 (charcoal), 150 mg/ml in water. Vortex 30 sec.

Centrifuge out charcoal, 1500 rpm, 5 min.

Count supernatant plus 3 ml scintillation fluid in liquid scintillation counter.

20

**Assay #2.**

The following reagents are mixed together:

**Final concentration:**

25  $2 \times 10^6$ - $1 \times 10^7$  cpm [<sup>3</sup>H] GTP,

dried down in SpeedVac

0.1 mM GTP

10 mM MgCl<sub>2</sub>

50 mU inorganic pyrophosphatase

30 10 mM fucose 1-phosphate

50 mM MOPS or Tris-HCl, pH 7.5

1 mM GTP

12.5 mM KF

(phosphatase inhibitor, optional)

Enzyme or enzyme plus buffer for balance of volume, total reaction volume 20  $\mu$ l.

Incubate 5 min, 37° C.

5 Add 10  $\mu$ l 1 M ammonium sulfate. Mix.

Add 200  $\mu$ l ice cold methanol. Mix.

Microfuge 2 min.

Count 100  $\mu$ l supernatant (or a smaller amount if desired) plus 3 ml scintillation fluid in liquid scintillation counter.

10 Place remaining supernatant in fresh tube. SpeedVac to dryness.

Resuspend in 5  $\mu$ l ethanol. Spot 2  $\mu$ l on an aluminum-backed silica TLC plate. Use a 1/10 dilution of NEN [ $^{14}$ C]-GDP-fucose as standard. Dry with low heat.

Run in 7:3 Ethanol/1 M ammonium acetate pH 7.5, 2-3 hours or till solvent front was at least halfway up the plate. Air dry.

15 Spray with En $^3$ Hance. Let dry in hood for 10 min. Place in cassette with film (e.g., Kodak XAR film) and an intensifying screen, store at -70° C for 2h or overnight if needed.

Preparation of samples for assay: When transfected cells were assayed, 20 only assay #2 was used. The cells (from two six well plates) were washed twice with 2 ml/well calcium-magnesium-free PBS and then removed from the plates in the same buffer (2 ml/well) with a cell scraper. The cells were centrifuged in a Beckman GPKR centrifuge, 1500 rpm, for 15 min. The pellet was washed with 2 ml of the PBS and centrifuged as before. Each sample was resuspended in 100  $\mu$ l of lysis buffer (50 mM Tris-HCl, pH 7.5; 1% (v/v) Triton X100; 10 mM KF; 0.5 mM AEBSF; and 2  $\mu$ g/ml each antipain, aprotinin, chymostatin, leupeptin and pepstatin). The samples were 25 homogenized with a mini-dounce.

#### Assay #3. HPLC Assay of GDP-Fucose Pyrophosphorylase (GDPFPP)

30 A quantitative reverse phase ion pairing HPLC assay for GDPFPP from whole cell lysates has been developed that resolves GDP-fucose from GDP-mannose as well as GTP, GDP, GMP and guanosine. The elution buffers providing the best resolution were 20 mM potassium phosphate (monobasic), 10 mM tetrapropylammonium

phosphate, pH 5.0 with  $H_3PO_4$  (buffer A) and methanol (buffer B), using a linear gradient of buffer B from 1% to 20% over 20 minutes. Cells expressing recombinant GDPFPP were lysed with nonionic detergent and assayed in the presence of 50 mM Tris, pH 7.5; 10 mM magnesium chloride; 10 mM [ $^3H$ ]GTP; 10 mM fucose-1-phosphate and 5 10 mUnits of pyrophosphatase for five minutes at 37° C. Assays were quenched with methanol, pelleted, the supernatant was dried under vacuum and resuspended in water for analysis by reverse phase ion pairing HPLC. The formation of GDP-fucose was followed by absorbance and fractions were collected for scintillation counting. Since the specific activity of [ $^3H$ ]GTP in each assay is known, the CPM corresponding to GDP-fucose 10 eluting off the HPLC column can be quantitated. Using this assay, 1 ml of recombinant baculovirus GDPFPP yielded 35 mUnits where a unit is defined as 1  $\mu$ mole of GDP-fucose formed per minute. Although linearity of this assay has yet to be established, under the conditions utilized, less than 10% of the GTP provided to the cell lysate was consumed, indicating that the assay is in the linear range.

15

### 3. Photolabeling.

$N_3-[^{32}P]$ -GDP-fucose was synthesized as described for  $N_3-[^{32}P]$ -GDP-mannose in Szumilo, *et al.*, *J. Biol. Chem.*, 268, 17943-17950 (1993), except that partially purified GDPFPP and fucose 1-P were used. The photolabeling 20 experiments were carried out essentially as described in Szumilo, *et al.*, *J. Biol. Chem.*, 268, 17943-17950 (1993) and Potter, *et al.*, *Meth. Enzymol.*, 91, 613-633 (1983).

Highly purified enzyme (10  $\mu$ l) was incubated with approximately 100  $\mu$ M  $N_3-[^{32}P]$ -GDP-fucose at room temperature and exposed to short-wave UV radiation via a hand-held UV lamp at a distance of 1.5 cm for 1 min. Novex SDS-PAGE sample buffer 25 containing 1% BME was added to the samples and they were subjected to electrophoresis on 8% or 8-16% gradient gels. Autoradiography was performed on dried gels.

Using the above techniques, a 66,000 MW protein that correlated with GDPFPP activity and photolabeled with  $N_3[^{32}P]$ -GDP-fucose was identified.

### 30 4. Blotting, Proteolysis and Sequencing.

**Preparation of Enzyme for Sequencing:** After chromatofocusing, the pooled enzyme sample (fractions 18-22, 18 ml) was concentrated to 350  $\mu$ l with an

Amicon concentrator, with a Diaflo membrane. A 200  $\mu$ l aliquot of that material was concentrated by lyophilization and resuspended in 50  $\mu$ l DI water.

5                   **Blotting:** To 40  $\mu$ l of the lyophilized, resuspended sample, 40  $\mu$ l Novex SDS-PAGE sample buffer was added, along with 10 mM each DTT and mercaptoacetic acid. The mixture was heated at 100° C for 5 min. The sample was subjected to electrophoresis in a Novex 8% gel with 10 mM mercaptoacetic acid in the top chamber of the gel box. After electrophoresis, the gel was equilibrated in 10 mM CAPS, pH 11.0 with 10% methanol for 15 min. The proteins were blotted onto ABI PVDF (ProBlott) for 1.5 h at 20 V. The blot was washed with water, stained with 0.2% Ponceau S in 1% acetic acid destained in 1% acetic acid and washed with water. The appropriate protein was excised from the blot, placed in a 1.5 ml centrifuge tube, washed with water, and kept moist. The sample was stored at -20° C until shipping.

10

15                   **Proteolysis and Sequencing:** The isolated protein was subjected to *in situ* proteolysis with trypsin. Before pooling and proteolysis, 10% of each sample was subjected to amino acid analysis. Peptides were separated on a Vydac C-18 column. Strong, symmetrical peaks were chosen for sequence analysis.

### Example 2

#### PCR amplification, Isolation and Cloning of cDNAs Coding for the Enzyme

20                   Using the amino acid sequences of portions of the porcine protein, human nucleic acid sequences comprising sequences that could encode these partial amino acid sequences were identified in computer databases comprising expressed sequence tags (ESTs). A full-length cDNA was then obtained by polymerase chain reaction using the following primers and amplification conditions.

25                   The first human EST that matched with a porcine GDPFPP peptide sequence was GenBank accession # T75166 (NCBI gi: 691928). This EST is from the Merck-Washington University sequencing project and is 465 basepairs long, 335 bp of which was considered high quality sequence. A shorter identical EST sequence was also identified in the search, accession # F12805 from Genethon's cDNA sequencing project.

30                   Another two ESTs (not shown) were discovered after large sequences of the gene had been isolated and sequenced.

The full-length cDNA was cloned into an appropriate vector and expressed in COS cells as follows.

**1. Source of DNA template for PCR amplification.**

A cDNA library was made from the an Epstein-Barr-virus-transformed B lymphoblastoid cell line, JY (Terhost, C., *et al.*, *Proc. Natl. Acad. Sci. USA* 73:910), into the plasmid DNA vector pBSIISK(+) (Stratagene) following the protocols for RNA isolation, mRNA purification and cDNA library construction found in **GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL**, M. Kriegler, W.H. Freeman and Company (1990). A 0.1 ml aliquot of the library glycerol stock was inoculated into 500 ml LB growth media, supplemented with ampicillin at 50  $\mu$ g/ml, and grown at 37° C for 16 hours with agitation. The library DNA was isolated using the Qiagen Plasmid Isolation Kit following conditions provided by the vendor.

**2. PCR amplification.**

The human GDPFPP cDNA was amplified from the JY-cDNA library using native Pfu polymerase (Stratagene) and the polymerase chain reaction (PCR) using Buffer #1 and conditions supplied by the manufacturer. The primers generate Eco RV restriction sites at the 5' and 3' termini. The 5' and 3' primers were 5'-TCA-GAT-ATC-GGG-GCT-ATG-GCA-GCT-GCT-AG-3' and 5'-ATA-GAT-ATC-TCT-GGA-ATG-TTA-CTC-AAA-AAG-GCA-A-3', respectively. The PCR procedure amplified a 1.8 kb fragment.

20

**3. Cloning the GDPFPP cDNA into expression vectors.**

The amplified GDPFPP gene was digested with the enzyme Eco RV (New England Biolabs) using conditions provided by the manufacturer. The blunt-end 1.8 kb fragment was fractionated on an agarose gel and isolated using standard techniques. The 1.8 kb fragment was cloned into the Eco RV site of pcDNA3 (Invitrogen), the Sma I site of pEE12 (Celltech), and the Sma I site of pVL-1392 (Invitrogen) following standard techniques, generating pcDNA-GFP6, pEE12-GFP7, and pVL1392-GFP6, respectively.

**4. Transient expression of GDPFPP in COS cells.**

COS cells were transfected via deae-dextran with either plasmid pcDNA3 or plasmid pcDNA-GFP6 following the protocol in **GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL**, M. Kriegler. After a 72 hour incubation, cell lysates were

assayed for GDPFPP activity by the TLC-GDPFPP assay. Cells transfected with pcDNA-GFP6 generated GDP-fucose from fucose-1-phosphate and GTP.

5. **Stable expression of GDPFPP in NSO cells.**

NSO cells were transfected by electroporation with plasmid pEE12-GFP7 and glutamine synthetase positive clones were isolated following the procedures in [PATENT: NS0-GT]. Clones were screened for GDPFPP activity by the TLC-GDPFPP assay. Clone GFP-5 was assayed via the HPLC-GDPFPP to produce 1.9 nanounits GDPFPP/cell.

10

6. **Construction of GDPFPP-recombinant baculovirus and GDPFPP expression.**

Stock cultures of *Spodoptera frugiperda* insect cells (Sf9II; Gibco/BRL) were co-transfected with pVL1392-GFP6 and linearized BaculoGold viral DNA (PharMingen) using the transfection protocol of Invitrogen Inc. Recombinant GDPFPP-baculoviruses were isolated and a high-titer stock generated by standard techniques. Sf9II cells were infected at a multiplicity of infection of 3. Following a 72-hour incubation at 27° C the cells were harvested and GDPFPP purified. GDPFPP was assayed via the HPLC-GDPFPP to produce 5.8 nanounits GDPFPP/cell.

20

**Example 3**

**Assay of the Transfected Cells**

25 Lysates of COS cells transiently transfected with vector alone or with vector containing the full length cDNA clone were assayed for GDPFPP activity using Assay #2 described in Example 1. Each sample was assayed with either fucose 1-phosphate or an inappropriate sugar phosphate, glucose 1-phosphate. Specific GDPFPP activity was observed in the COS cells expressing the cDNA. In particular, GDP-fucose was detected in assay products of lysates from the cells transfected with the GDPFPP clone, but not in the vector alone condition. Additionally, a significant quantity of nucleotide sugar was not synthesized when glucose 1-phosphate was offered 30 as a substrate.

**Example 4****Recombinant GDPFPP from Baculovirus-infected SF9 Cells  
Can Be Used in the Fucosyltransferase Cycle  
for the Production of Cylexin™**

5

Cylexin™ is the brand name for a pentasaccharide that inhibits the adhesion of cells that express NeuAc $\alpha$ 2 $\rightarrow$ 3Gal $\beta$ 1 $\rightarrow$ 4(Fuc $\alpha$ 1 $\rightarrow$ 3)GlcNAc (SLe $X$ ) to selectin receptors such as ELAM-1. Recombinant GDPFPP from baculovirus-infected SF9 cells can be used in the fucosyltransferase cycle for the production of Cylexin™.

10

**1. Synthesis of GFP in Baculovirus cultures.**

SF-9 cells were grown and passaged in SF-900 II medium (GIBCO-BRL) in 50 ml shaker flask culture, under conditions recommended by the media manufacturer. The cells were grown at 27°C in the dark with continuous shaking at 130 rpm. To create a recombinant virus, standard protocols were used (see O'Reilly DR, Miller LK, and Luckow V (1992), **BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL**. New York; W.H. Freeman Co.) with some modifications. Cells were co-transfected with Baculogold baculovirus DNA [Pharmingen, San Diego, CA] and the purified transfer vector, according to the Baculogold manufacturers directions. Transfected cells were incubated in SF-900 II medium for five days after transfection, when the supernatant was collected. Recombinant virus was isolated by limiting dilution on fresh SF-9 cells in 60-well tissue culture dishes [Robbins Scientific].

After 7 days, virus was collected from wells containing infected cells at dilutions that yielded less than 20% infected wells. Twelve recombinant viruses were expanded in volume by infection of fresh cells in 24-well tissue culture dishes. After 7 days, each virus isolate was screened for its ability to synthesize GFP, and each isolate was found to be positive. One isolate (designated E-9) was chosen for further expansion by infection of fresh SF-9 cells in a 225 cm<sup>2</sup> flask. The virus stock produced after 7 days of infection was collected and stored as virus master stock, designated rBacy GFP p3.

SF-9 cells were grown to 1000 ml shake flask culture in a 2800-ml Fernbach flask in suspension culture at 27°C, 130 rpm until the culture reached a density of 8 x 10<sup>6</sup> cells per ml of culture. Virus titre of the master stock was assumed to be 3 x 10<sup>8</sup> pfu/ml, and the SF-9 cells were infected at a multiplicity of infection of

approximately 0.2 pfu/cell in the following manner: 5 ml of virus stock were added to the cells in the Fernbach flask. The flask was swirled by hand 4-5 revolutions to distribute the virus. 500 ml of fresh media were added, and the flask was returned to the shaking incubator for 48 hours. After incubation, the virus was collected aseptically in the supernatant. The culture was transferred to sterilized one-liter bottles and the cells were centrifuged away (3,000 x g, 15 min) from the supernatant (virus working stock), which was collected and stored. The titre of the virus working stock was determined by calculation from the results of limiting dilution infection, as described in O'Reilly et al., and was determined to be  $2.4 \times 10^9$  pfu/ml.

To produce GFP, Sf-9 cells were grown to 1000 ml shake flask culture in three 2800-ml Fernbach flasks in suspension culture at 27 C, 130 rpm until the cultures reached a density of approximately  $8 \times 10^6$  cells per ml of culture. The cells were infected at a multiplicity of infection of approximately 3 pfu/cell using the method described above, with the modification of addition of 10 ml of virus working stock to each flask to achieve the desired multiplicity. After addition of virus, 500 ml of fresh media was added to each flask, and the cultures were returned to the shaking incubator for 65 hours. A one-ml analytical sample was collected and centrifuged (5000 x g, 2 min), and the cells and supernatant were analyzed for GFP content. The remaining cells were harvested by centrifugation in one-liter bottles (3,000 x g, 25 min), resuspended in a small amount of fresh medium, combined, and recentrifuged. Supernatants were discarded.

## 2. Partial Purification of GDPFPP from Baculovirus-infected Sf9 Cells.

A 75 x g cell pellet was obtained from centrifugation of a 4 L culture of Sf9 cells infected with baculovirus harboring the cDNA for GDPFPP. The pellet was resuspended in 50 Mm Tris-HCl, pH 7.8 plus 2  $\mu$ g/ml each antipain, aprotinin, chymostatin, leupeptin and pepstatin. The sample was sonicated with a Tekmar Sonic Disruptor at 60% power for 6-15 sec. bursts. The sample was centrifuged at 4° C in a Sorvall RC-5B centrifuge in a GSA rotor at 8000 rpm for 30 min. The pellet was discarded. While the supernatant fraction was stirring rapidly, 3.5 ml of a 10% (v/v) solution of polyethylen(imine) was added. The mixture was centrifuged at 4° C in a Sorvall RC-5B centrifuge in a GSA rotor at 8000 rpm for 15 min. The pellet was discarded. Solid ammonium sulfate was added to 60% saturation and the sample stirred

at 4° C for 1 hour. The mixture was centrifuged at 4° C in a Sorvall RC-5B centrifuge in a GSA rotor at 8000 rpm for 15 min. The supernatant fraction was discarded. The pellet was resuspended in 20 ml 50 mM Tris-HCl, pH 7.8.

5       3.     The Fucosyltransferase Cycle.

A fucosyltransferase cycle was run in a volume of 100  $\mu$ l with the following reagents: 50 mM sialy-N-acetyllactosamine; 50 mM HEPES, pH 7.5; 100 mM phosph(enol)pyruvate, adjusted to pH 7.5; 2 mM guanosine 5'-diphosphate; 0.7 U pyruvate kinase; 50 mM fucose 1-phosphate; 5 mM MgCl<sub>2</sub>; 0.5 mg/ml (w/v) bovine serum albumin; 0.2% (w/v) NaN<sub>3</sub>; 15% (v/v) GDPFPP purified as described above; and 15% (v/v) fucosyltransferase V ( $\alpha$ -1,3 fucosyltransferase, 2.16 U/ml, partially purified on SP-Sepharose). The cycle was incubated at 37° C overnight. A 0.3  $\mu$ l aliquot of a 10 mM solution was clearly apparent in the lane containing the fucosyltransferase cycle sample. A 0.3  $\mu$ l sample of the cycle spiked with 0.3  $\mu$ l of a 3.3 mM solution of 15 Cylexin™ standard was run on the same plate to confirm the identity of the cycle product as Cylexin™.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

Sequence I.D. No. 1  
 Sequence Range: 1 to 2318

10 20 30 40 50 60  
 CGGTGCTTC CGGGCGGCTC TCAGCCAGG TGGGCTATC GCAGCTGCTA GGGACCCCTCC  
 CGCACGACAC CCCGGCCAG AGTCCCTTCC ACCCCGATAC CCTCGACGAT CCCGGAGG  
  
 70 80 90 100 110 120  
 GGAAGTATCG CTGGGAGAAG CCACCCAGGG AAAATTGGGG AGCTTTTCCC AGCTAAGAGG  
 CCTTCATACC GACGCTCTTC GGTGGCTGG TTTAACGCC TCCAAAGCC TCGATTCTCC  
  
 130 140 150 160 170 180  
 CAAACTTGTG GCACGTGGAG AATTCTGGG CATACTTCCA ATAACAGGG CTGATGAAA  
 GTTTCACAT CGTGCACCTC TTAAGACCC CTATCACCT TATTCTGCC GACTACTTT  
  
 190 200 210 220 230 240  
 ACAGGAACCT GTTACAAACC AACACCTGTC AGAAAAGCTG AAAAGAAAGG ACTTACCCCT  
 TGTCTTGA CGAATGTGG TTGTCGACAG TCTTTCCAC TTTCTTCCC TCAATGGGA  
  
 250 260 270 280 290 300  
 TGGACTTCAA TATCACCTT TTGTCGATCC TCTGGACCC AAAATTGGAA ATGGAGGATC  
 ACCTCAACTT ATAGTGCAAA AACACCTAGG ACCACCTGG TTTAACCTT TACCTCCTAG  
  
 310 320 330 340 350 360  
 AACACTTGT GCCTTCAAT GTTGGAAAA CCTATATGGA CATAATGCA ATTCTTTTAC  
 TGTGAAACA CGGAAAGTTA CAAACCTTTT CGATATACTT CTATTTACCT TAAGAAAATC  
  
 370 380 390 400 410 420  
 CATCTTATTA ATTCACTCTC CTGGCTACAG TCAACGACTT CCAAATGCA CTGCTCTGG  
 CTAGAATAAT TAAGTGAGAC CACCGATGTC ACTTGGTGA GTTTACGTT CACGAGACCC  
  
 430 440 450 460 470 480  
 AAAATTTTC ACTGCTTAC CTCTTGGTAA CCCCATTTAT CAGATGCTAG AATTAAAGCT  
 TTTTAAAG TGACGAAATG GAGAACCTT CGCTAAATA GTCTACGATC TTAATTTCGA  
  
 490 500 510 520 530 540  
 ACCCATGTC ATTGATTTCC CCTTAAATAT GAATCTGGG ATTCTGGTTA CCTGGCAGA  
 TCGGTACATC TAACTAAACG GGAATTTATA CTTAGGACCT TAAGACCAAT GGACACCTCT  
  
 550 560 570 580 590 600  
 TGATATTGAA CTTATAGTA TTGGACAATT TGAGTTTATT AGCTTGCACA AACCTGGCTT  
 ACTATAACTT GAAATATCAT AACCTCTTAA ACTCAAATAA TCCAAACTGT TTGGACCGAA

610            620            630            640            650            660  
 TACTGCTTTA GCTCATCCTT CTAGTTGAC GATAGGTACC ACACATGGAG TATTTGCTT  
 ATGACGAAAT CGAGTAGGAA GATCAAGTC CTATCCATGG TGTGTACCTC ATAAACAGAA  
  
 670            680            690            700            710            720  
 AGATCCTTTT GATGATTTAA AACATAGAGA CCTTCAATAC AGGTCTTGCCT ATCGTTTCT  
 TCTACGAAAAA CTACTAATT TTCTATCTCT CGAACTTATG TCCAGAACGG TAGCAAAGGA  
  
 730            740            750            760            770            780  
 TCATAAGCCC ACCATAGAAA AGATGTATCA CTAAATGCT GTGTGTAGAC CTGGAAATTT  
 AGTATTCGGG TCGTATCTT TCTACATAGT CAAATTACGA CACACATCTG GACCTTTAAA  
  
 790            800            810            820            830            840  
 TTGTCAACAG GACTTTCCTG GGGGTGACAT TCCCCATCTT AAATTAGACT CTGACTATGT  
 AACACTTGTCTC CTGAAACGAC CCCCACCTCA ACGGCTAGAA TTTAATCTGA CACTCATACA  
  
 850            860            870            880            890            900  
 CTACACACAT AGCCTATTTT ATATGGATCA TAAATCAGCA AAAATGTTAC TTGCTTTTTA  
 GATGTGTCTA TCGGATAAAA TATACTAGT ATTTAGTCGT TTTTACAATG AACGAAAAAT  
  
 910            920            930            940            950            960  
 TGAAAAAAATA GGCACACGTGA GCTGTGAAAT AGATCCCTAT GGTGACTTTG TCCAGGCCTT  
 ACTTTTTTAT CGGTGTCACT CGACACTTTA TCTACGGATA CCACTGAAAG ACGTCCGAAA  
 Q A L>
  
 970            980            990            1000            1010            1020  
 GGGACCTGGA CCAACTGTCC AGTACACCGAG AAACACATCA CATGTCATTA AAGAAGAGTC  
 CCCTGGACCT CGTTGACACC TCATGTGGTC TTGCTGTAGT GTACAGTAAT TTCTTCAG  
  
 1030            1040            1050            1060            1070            1080  
 AGAGTTGGTA GAAATGAGGC AGAGAATATT TCATCTTCTT AAAGGAACAT CACTAAATGT  
 TCTCAACCAT CTTTACTCCG TCTCTTATAA ACTAGAAGAA TTTCCTTGTGTA GTGATTACA  
  
 1090            1100            1110            1120            1130            1140  
 TGTGTGTTCTT AATAACTCCA AATTTTATCA CATTGGAACA ACCGAAGAAAT ATTTGTTTTA  
 ACAACAAAGAA TTATGAGGT TTTAAATAGT GTAACCTTGT TGGCTTCTTA TAAACAAAAT

1150            1160            1170            1180            1190            1200  
 CTTTACCTCA GATAACAGTT TAAAGTCAGA CCTCCCCCTTA CAGTCCATAA CTTTGTAGTAT  
 CAAATCGAGT CTATTTGTCAA ATTTCACTCT CGAGCCGAAT GTCAGGTATT CAAACATCA

1210            1220            1230            1240            1250            1260

CTTTCAGAT ATACCAGAAT GCTCTGGCAA AACATCCCT ATCATTCAAA GCATACTGCA  
CAAAGGTCTA TATGGTCTTA CGAGACCGTT TTGTAGGACA TAGTAAGTTT CGTATGACCT

1270 1280 1290 1300 1310 1320  
TTCAGATGT TCTCTGGCAC CTGGCTCACT TCTGGAGTAT TCCAGATTCG GGCCTGATGT  
AAGTCTACA AGACACCGTG GACCGAGTCA ACACCTCATA AGTCTAACCC CGCGACTACA

1330 1340 1350 1360 1370 1380  
TTCAGTTGGG GAAAATGCA TTATTAGTGG TTCTTACATC CTAACAAAAAG CTCCCCCTCCC  
AAGTCACCC CTTTACGCT AATAATCACC AAGAATGTAG GATTGTTTTC GACGGGAGCC

1390 1400 1410 1420 1430 1440  
CCGACATTCT TTGTATGTT CCTTAAGCTT AAAGATGAAT AGATGCTTAA AGTATGCAAC  
GGGTCTAAGA AAACATACAA GGAATTGAA TTCTACTTA TCTACGAATT TCATACGTTG

1450 1460 1470 1480 1490 1500  
TATGGCATTG CGAGTGCAAG ACAACTTGA AAAGACTGTC AAAACATTGT CAGATATAAA  
ATACCGTAAAC CCTCACGTTG TGTGAACCTT TTCTCACAC TTTGTAAACA GTCTATATTT

1510 1520 1530 1540 1550 1560  
GTTACTTCAA TTCTTGGAG TCTGTTCTCT GTCATGCTTA GATGTTGGGA ATCTTAAAGT  
CAATCAAGTT AAGAAACCTC AGACAAAGCA CAGTACCAAT CTACAAACCT TACAATTTC

1570 1580 1590 1600 1610 1620  
TACAGAGGAA CTGTTCTCTG GAAACAAGAC ATGTCTGAGT TTGTGGACTG CACGCATTT  
ATCTCTCTT GACAAGAGAC CATTCTCTG TACAGACTCA AACACCTGAC GTGGTAAAGA  
T E E L F S G N K T C L S L W T A R I F >

1630 1640 1650 1660 1670 1680  
CCCACTTGT TCTCTTTGA GTGACTCACT TATAACATCC CAAAGATGT TAAATGCTGT  
GGTCAAAACA AGAAGAAACT CACTGAGTCA ATATTGTAGG GATTCTACAA ATTACCGACA

1690 1700 1710 1720 1730 1740  
TAAGAACAAAG TCAGCATTCA GCTGAATAG CTATAAGTTG CTGTCCATTG AAGAAATGCT  
ATCTCTTTC ACTCGTAAGT CGGACTTATC GATATTCAAC GACAGGTAAC TTCTTACGA

1750 1760 1770 1780 1790 1800  
TATCTACAAA GATGTACAAG ATATGATAAC TTACAGGGAA CAAATTTTC TAGAAATCAG  
ATAGATGTGT CTACATCTTC TATACTATTG AATGTCCCTT GTTAAAAAG ATCTTACGTC

1810 1820 1830 1840 1850 1860  
TTTAAAAAGC AGTTGATGT AGAGATATTG TAAATATTGT ACACCTTCCC TTTTGACTA  
AAATTTTCG TCAAACTACA TCTCTATAAA ATTTATAACA TGTCAAACCC AAAAATCTCAT

1870	1880	1890	1900	1910	1920
ACATTCAGA	GATAGGTATT	TTGGTAGGC	TGTTTCACTG	AACTCAGTTA	ATCAAAACTG
TGTAAGGTCT	CTATCCATAA	AAACCATCCG	ACAAAGTGAC	TTCAGTCAAT	TACTTTGAC
1930	1940	1950	1960	1970	1980
TATTAACATA	ATTGTTGTAG	CATAATATTA	ATAGTGCAAA	AGTACATATA	AGTCATTTG
ATAATTGTAT	TAACAAACATC	GTATTATAAT	TATCACGTTT	TCATGTATAT	TCAGTAAAAC
1990	2000	2010	2020	2030	2040
ATGAAAAATA	TTCCAAGACT	AACCTCAGAA	AAGAGATACT	ATTTTGGATG	TGTATCACTA
TACTTTTAT	AAGGTTCTGA	TTCAACTCTT	TTCTCTATGA	TAAAACCTAC	ACATAGTCAT
2050	2060	2070	2080	2090	2100
TTTTGTTTT	TAATAATCAT	TGATTTGTGG	AGCATTGTTT	TTTCACATAA	TTAGTTTTAA
AAAAACAAAA	ATTATTACTA	ACTAAACACC	TCGTAACAAA	AAAGTGTATT	AATCAAAATT
2110	2120	2130	2140	2150	2160
AGGTAATTTT	CTAACGCATAC	CTTTCGAATT	TTTCCATCTT	TTTTGAGGCT	TTTGGTCCAG
TCCATTAAAA	GATTCCTATG	GAAACCTTAA	AAACGTAGAA	AAAACCTCGA	AAACCAGGTC
2170	2180	2190	2200	2210	2220
TGAAGTTCTA	AGTATTCACT	GGCACTTCTC	TCCTCAACTG	TAATTCTATT	TTTAATAATA
ACTTCAGAT	TCATAAGTGA	CCGTCAAGAG	AGGAGTTGAC	ATTAAGATAA	AAATTATTAT
2230	2240	2250	2260	2270	2280
AAAATGCCAT	ACTGTAGGGT	CTTCAGAGTA	GTGTAGGAAT	ACTGTACAAA	TACTTTTCA
TTTTACCGTA	TGACATCCC	GAAGTCTCAT	CACATCCTTA	TGACATCTT	ATGAAAAAGT
2290	2300	2310			
GAAACGAATC	CATACGTCAC	AAATTCACTC	ACTCCCCA		
CTTTGCTTAG	GTATCGACTG	TTTAAGTGAG	TCACGGGT		

## Sequence I.D. No. 2

M A A A R D P P&gt;

E V S L R E A T Q R K L R R F S E L R G&gt;

K L V A R G E F W D I V A I T A A D E K&gt;

Q E L A Y N Q Q L S E K L K R K E L P L&gt;

G V Q Y H V F V D P A G A K I G N G G S&gt;

T L C A L Q C L E K L Y G D K W N S F T&gt;

I L L I H S G G Y S Q R L P N A S A L G&gt;

K I F T A L P L G N P I Y Q M L E L K L&gt;

A M Y I D F P L N M N P G I L V T C A D&gt;

D I E L Y S I G E F E P I R F D K P G F&gt;

T A L A H P S S L T I G T T H G V F V L>

D P F D D L K H R D L E Y R S C H R F L>

H K P S I E K H Y Q F N A V C R P G N F>

C Q Q D F A C G D I A D L K L D S D Y V>

Y T D S L F Y M D H K S A K M L L A F Y>

E K I G T L S C E I D A Y G D F L Q A L>

G P G A T V E Y T R N T S H V I K E E S>

E L V E M R Q R I F H L L K G T S L N V>

V V L N N S K F Y H I G T T E E Y L F Y>

F T S D N S L K S E L G L Q S I T F S I>

F P D I P E C S G K T S C I I Q S I L D>

S R C S V A P G S V V E Y S R L G P D V>

S V G E N C I I S G S Y I L T K A A L P>

A H S F V C S L S L K M N R C L K Y A T>

M A F G V Q D N L K K S V K T L S D I K>

L L Q F F C V C F L S C L D V W N L K V>

T E E L F S C N K T C L S L W T A R I F>

P V C S S L S D S V I T S L K M L N A V>

K N K S A F S L N S Y K L L S I E E M L>

I Y K D V E D M I T Y R E Q I F L E I S>

L K S S L M >

WHAT IS CLAIMED IS:

1. An isolated nucleic acid which encodes GDP-fucose pyrophosphorylase (GDPFPP).
- 5 2. The isolated nucleic acid of claim 1, wherein the nucleic acid has a sequence as shown in SEQ. ID. No. 1.
- 10 3. The isolated nucleic acid of claim 1, wherein the nucleic acid encodes a GDP-fucose pyrophosphorylase having a sequence as shown in SEQ. ID. No. 2.
- 15 4. A nucleic acid of claim 1, in a recombinant DNA construct wherein said nucleic acid is operably linked to a promoter.
- 20 5. A recombinant DNA construct according to claim 4, wherein the promoter directs expression of the nucleic acid in mammalian cells.
6. A recombinant DNA construct according to claim 4, wherein the promoter directs expression of the nucleic acid in insect cells.
- 25 7. A recombinant DNA construct according to claim 4, wherein the promoter directs expression of the nucleic acid in fungal cells.
8. A composition comprising an isolated GDPFPP enzyme encoded by a nucleic acid of claim 1, wherein the isolated GDPFPP has a molecular weight of about 66 kD.
- 30 9. The composition of claim 8, wherein the GDPFPP enzyme has a sequence as shown in SEQ. ID. No. 2.
10. A nucleic acid of Claim 1, which is contained in a cell that is genetically engineered to contain said nucleic acid.

**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US97/05968

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : A61K 38/51; C12P 21/06; C12N 9/12, 1/20, 15/00; C07H 21/04.  
US CL : 424/94.5; 435/69.1,194, 252.3, 320.1; 536/23.2.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/94.5; 435/69.1,194, 252.3, 320.1; 536/23.2.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
Sigma Catalog.

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Extra Sheet.

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COATES, S. W. Subcellular Localization of Sugar Nucleotide Synthetases. J. Biol. Chem. October 1980. Vol. 255. No. 19. pages 9225-9229, see the entire article.	1-10
A	STILLER, R. Enzymatic Synthesis of beta-L-Fucose-1-phosphate and GDP-fucose. Liebigs Ann. Chem. 1992. Vol. 5. pages 467-471, see the entire article.	1-10

Further documents are listed in the continuation of Box C.

See patent family annex.

•	Special categories of cited documents:		
*A*	document defining the general state of the art which is not considered to be of particular relevance	*T*	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*E*	earlier document published on or after the international filing date	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*L*	document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*O*	document referring to an oral disclosure, i.e. exhibition or other means	*&*	document member of the same patent family
*P*	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
30 MAY 1997

Date of mailing of the international search report  
26 JUN 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Authorized officer

TEKCHAND SAIDHA

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US97/05968

**B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

STN: Files - medline, caplus, wpids, biosis, scisearch and biotechds.  
APS (US patents). Search Terms : GDP-fucose pyrophosphorylase or GDPFPP, gene or dna, GDP-fucose, in different combinations.